

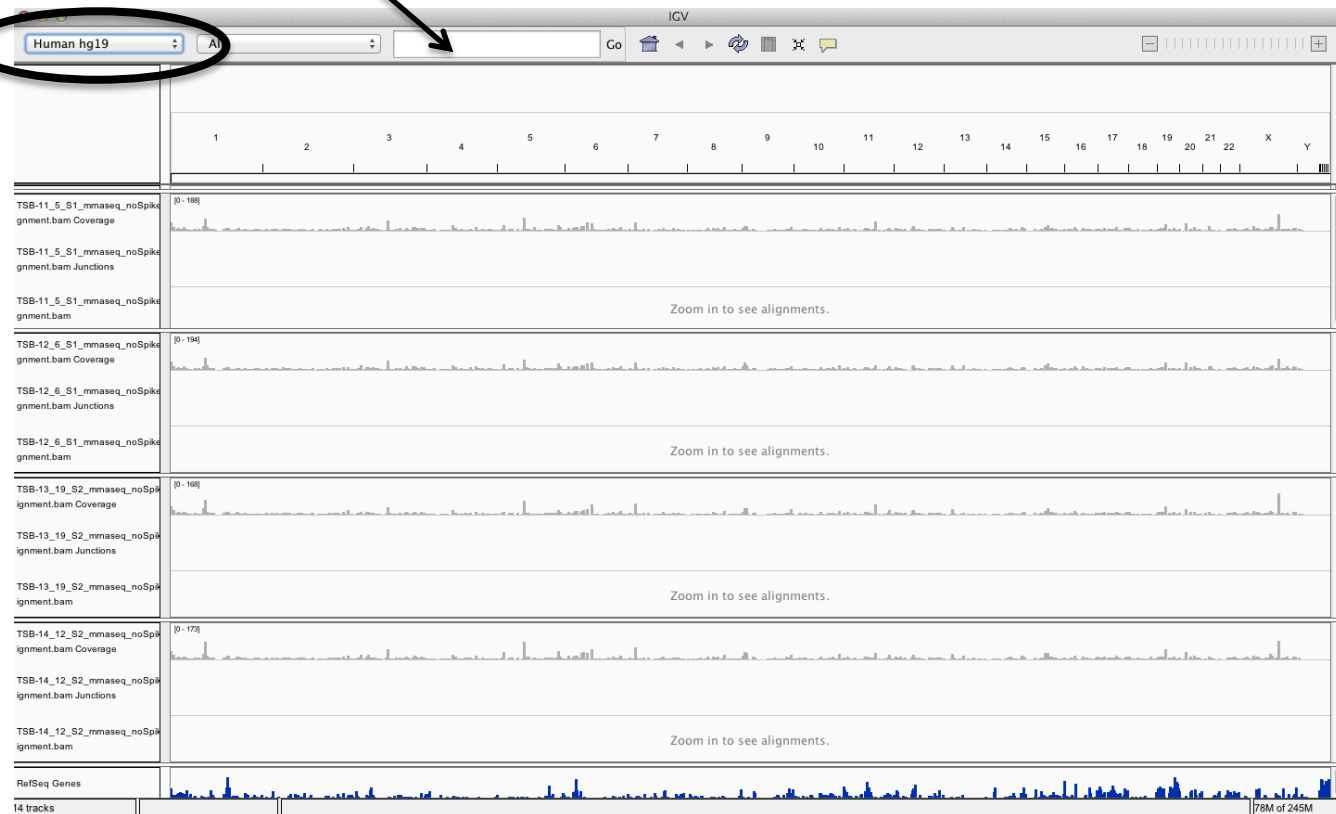


NGS read mapping : answers to questions

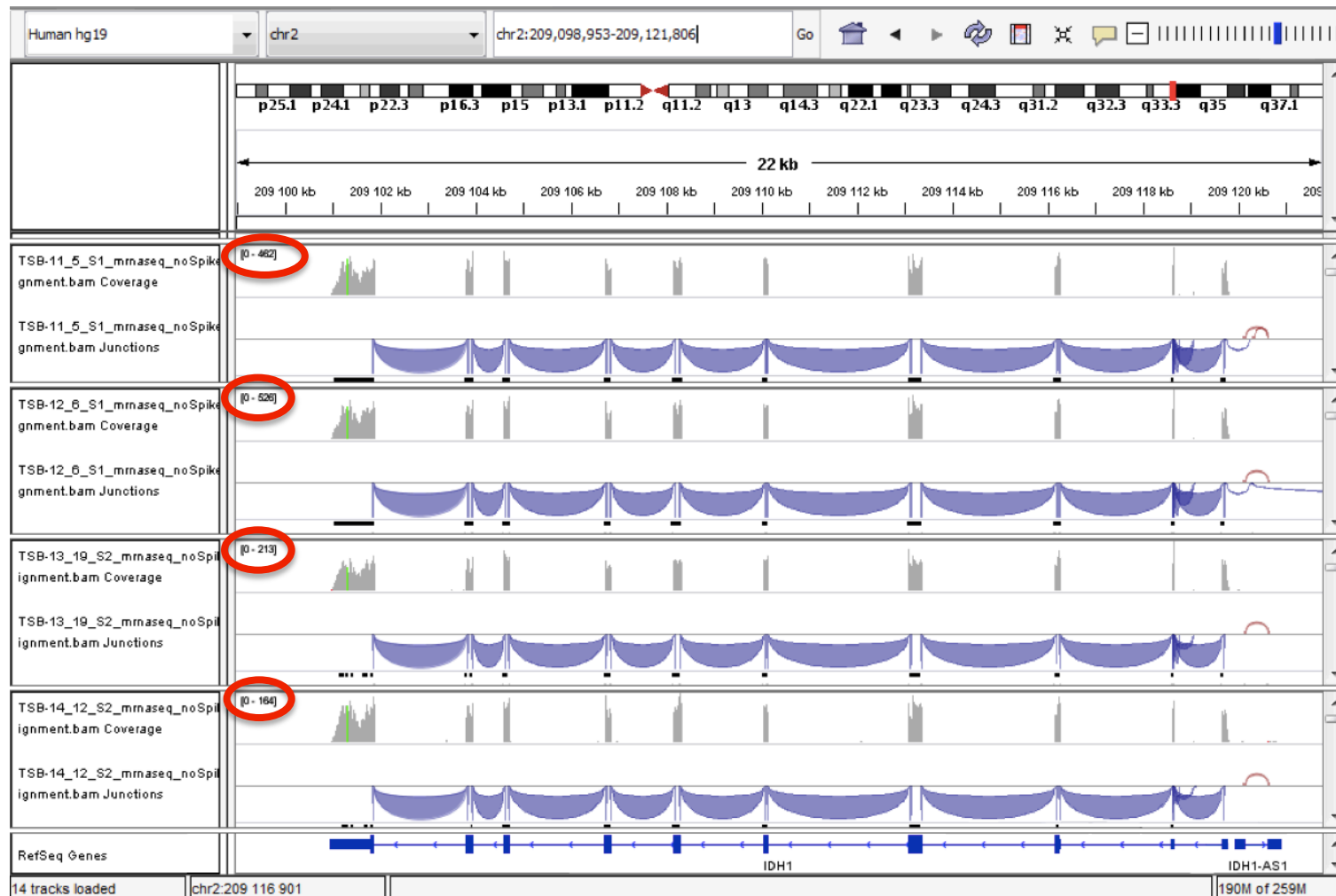
Céline Keime
keime@igbmc.fr

Question 1

- Genome assembly = hg19
- File → load from file and select the 4 BAM files
- Type IDH1 in the box and click Go

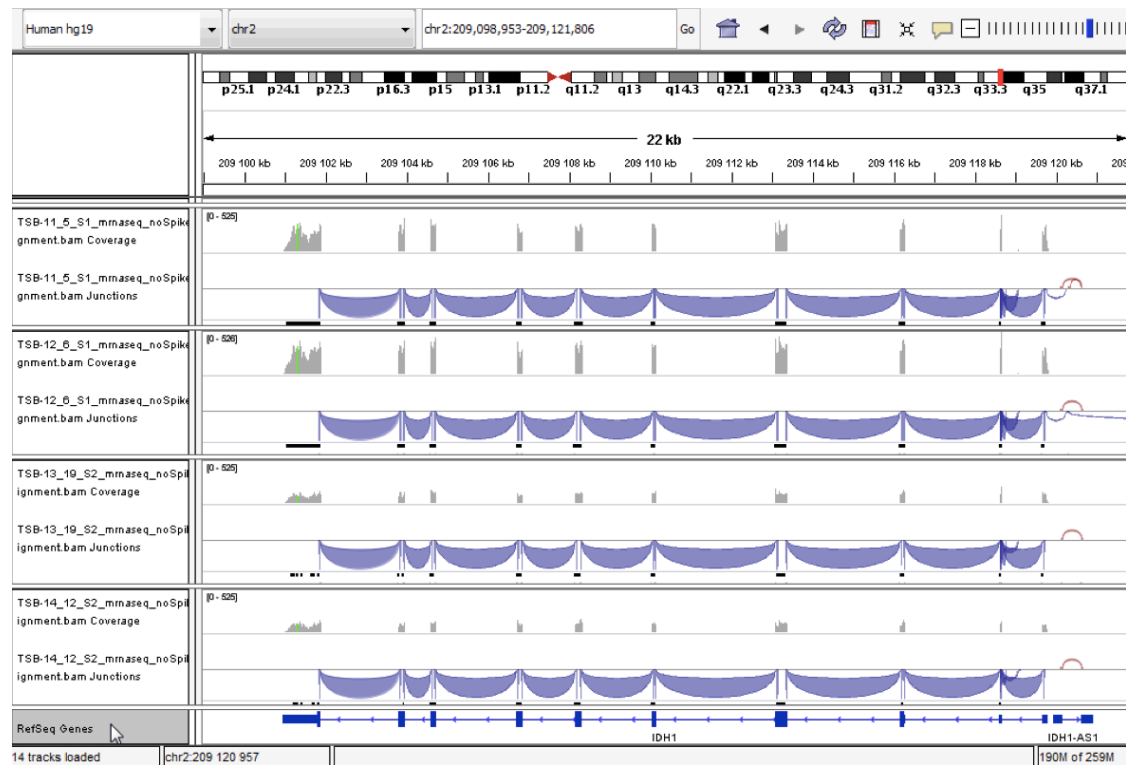


Question 1



Question 1

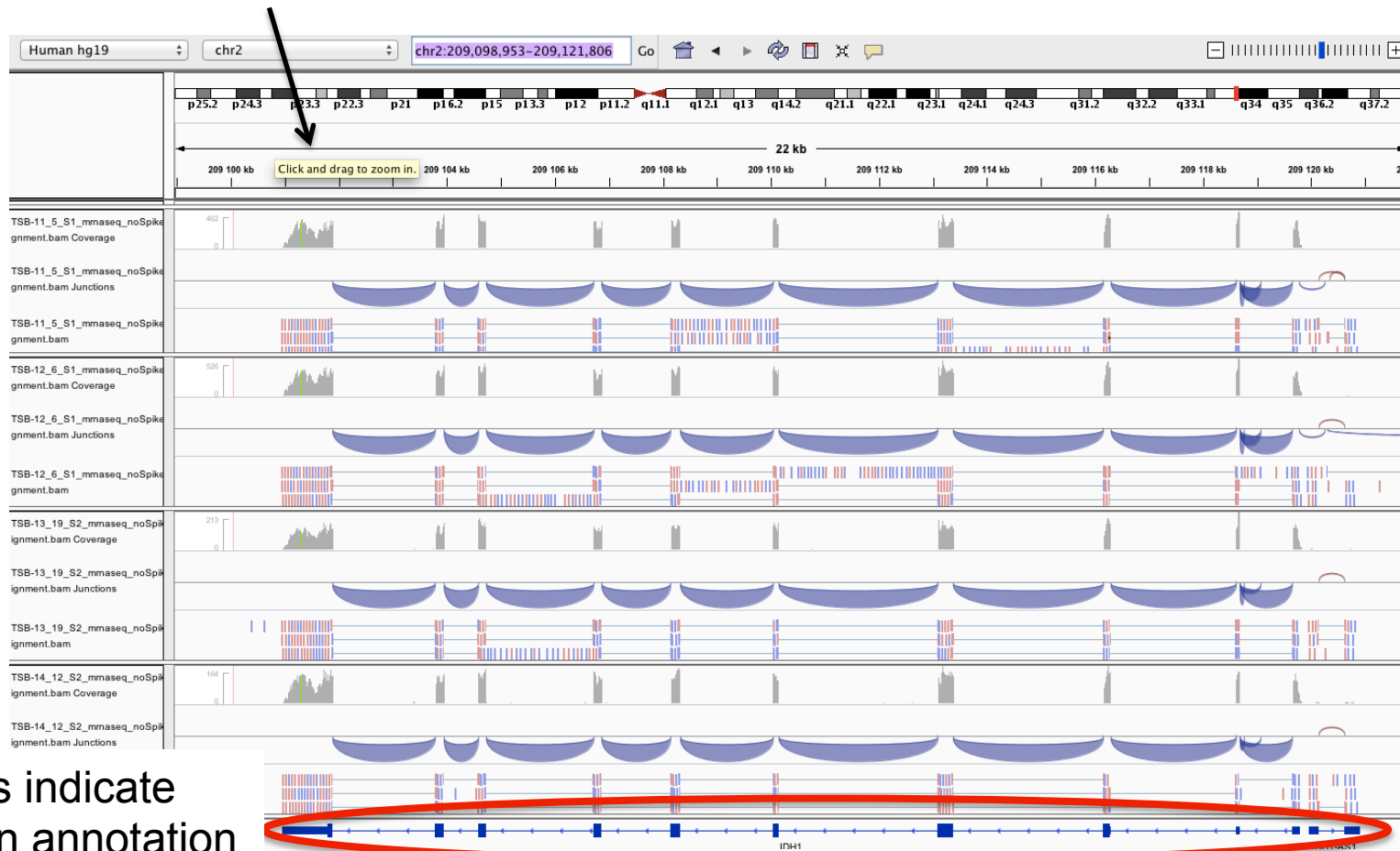
- Same scale for all tracks
 - Right-click on each coverage track → Set Data Range → Max = 525



- IDH1 is under-expressed in siMitf samples compared to siLuc ones
- More information on IDH1 → Click on gene name

Question 2

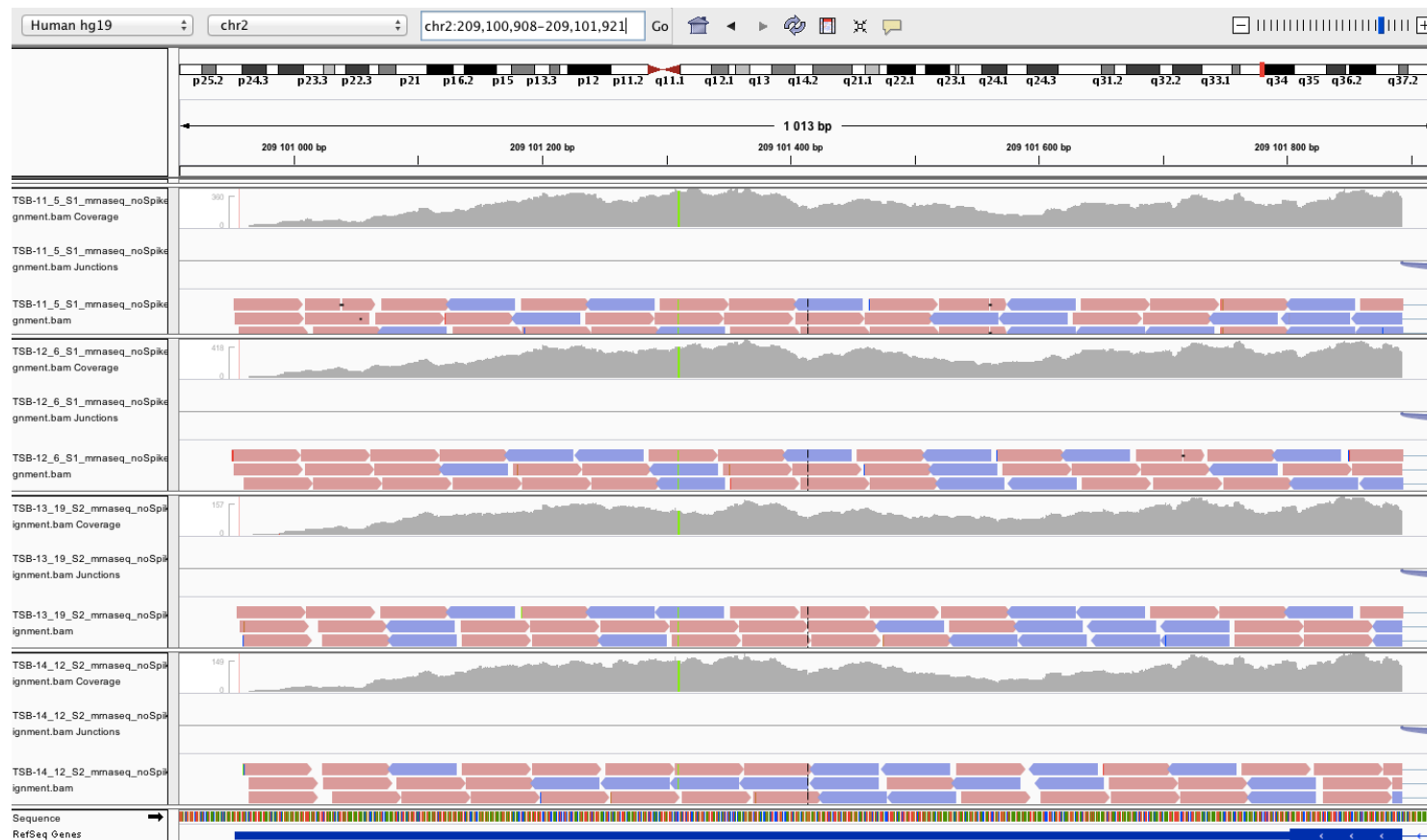
- Click and drag to define a window around the last exon in order to zoom in



Arrows indicate orientation annotation

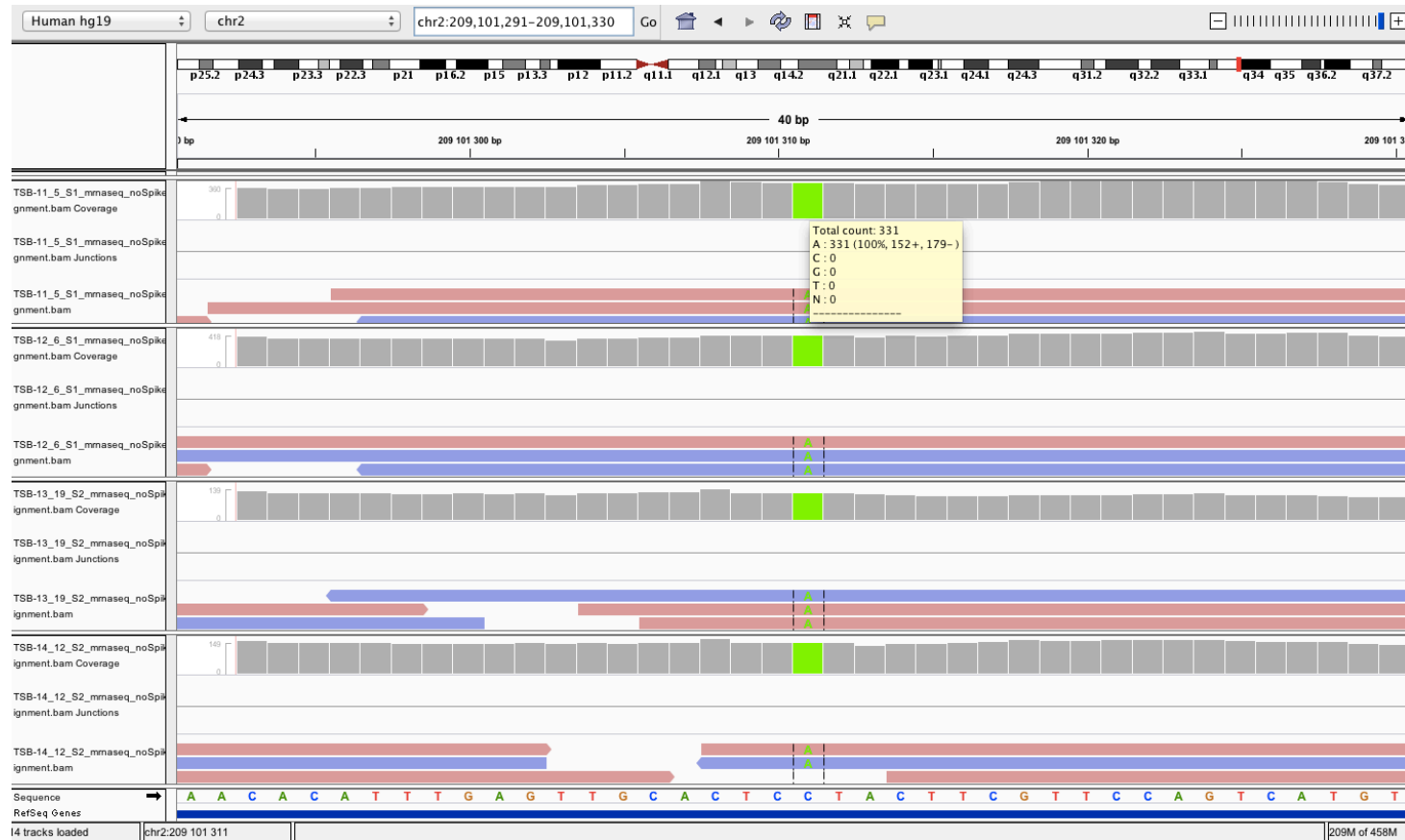
Question 2

- You can see the nucleotide difference in green
- Click and drag to zoom in around this position



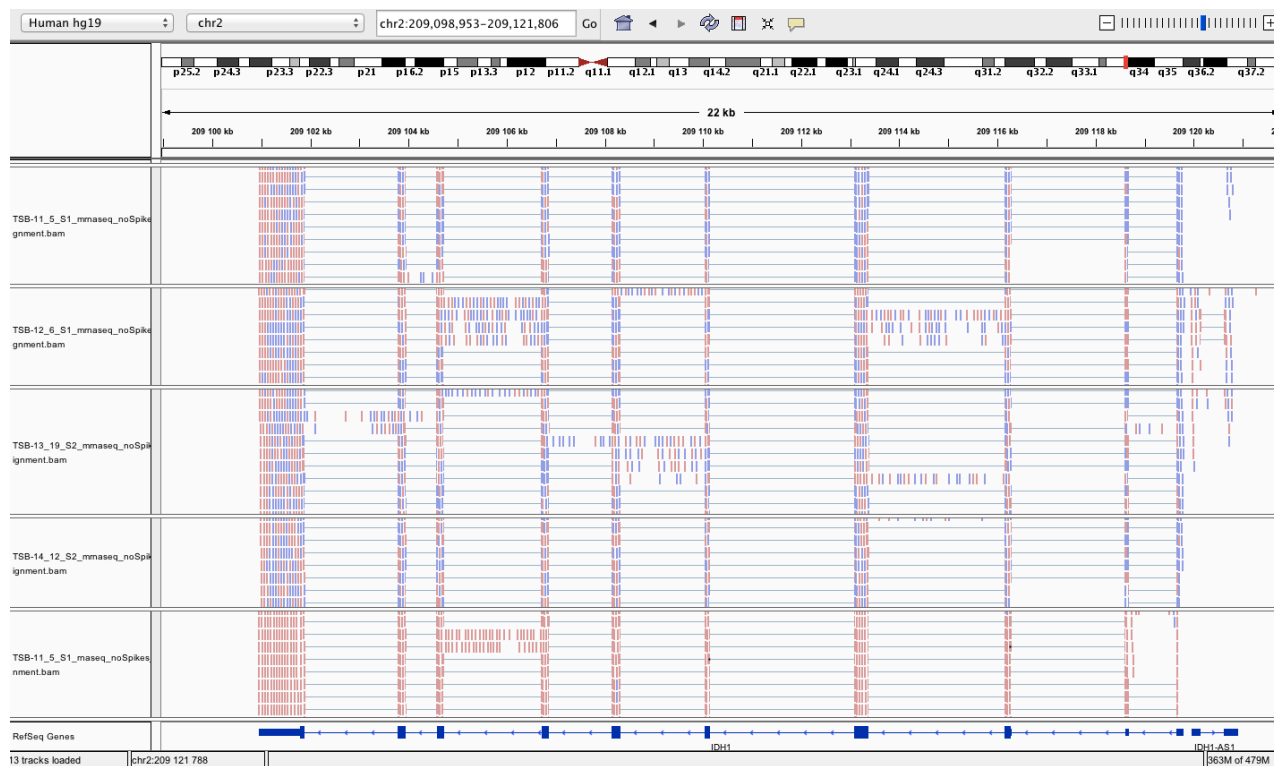
Question 2

- In the location chr2:209,101,311 :
 - A in 100% of the RNA-seq reads, C in the reference genome



Question 3

- File → load from file and select TSB-11_5_S1_rnaseq_noSpikes_alignment.bam
- For each BAM file choose Color alignments by → read strand
- e.g. IDH1 gene

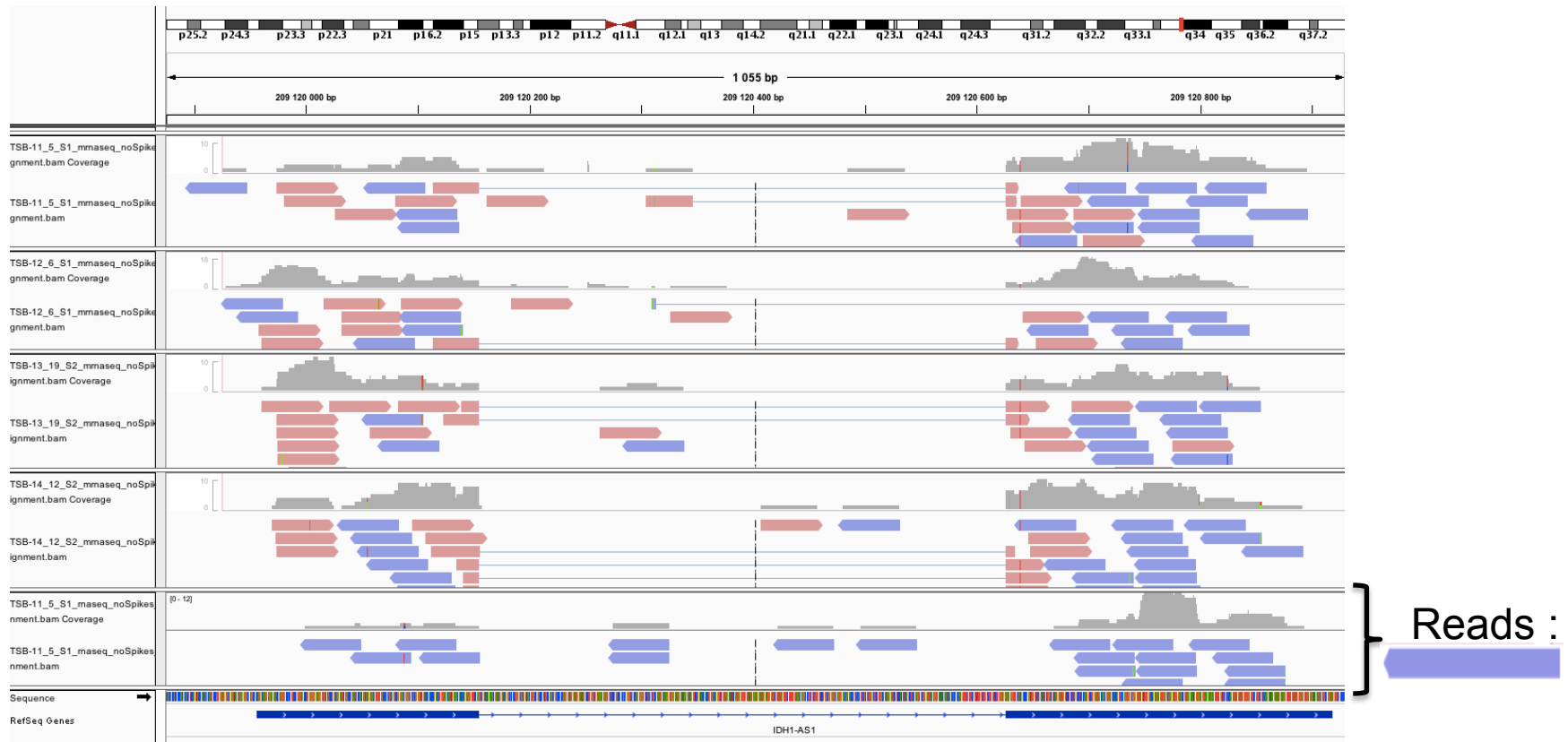


Reads :



Question 3

- e.g. IDH1-AS1 gene (IDH1 antisense RNA1)



- **The 2nd protocol is directional** (keep strand information)
 - All reads are in the opposite direction from the sense of transcription

Question 4 : additional information

■ Display of splice junctions

■ Strand

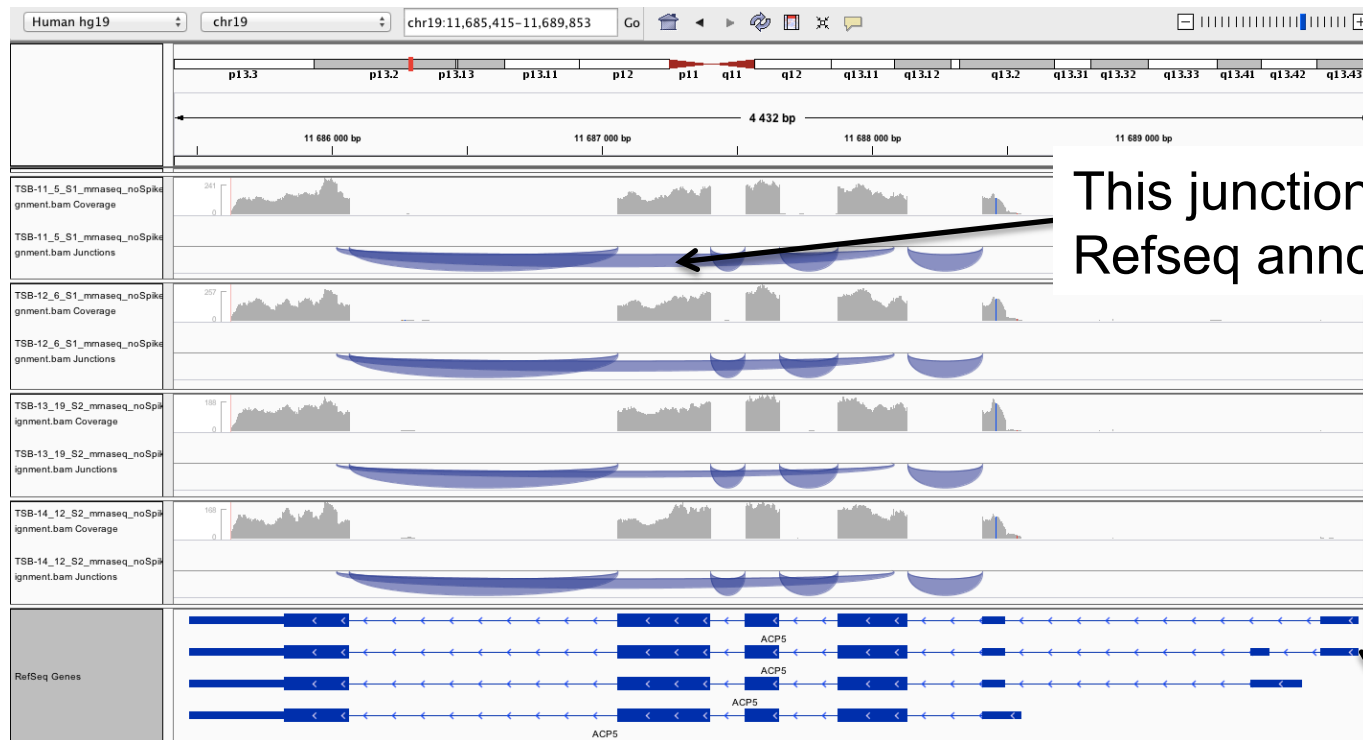
- Blue junctions : + strand
- Red junctions : - strand

■ Depth of coverage

- The thickness of the arcs are proportional to the depth of coverage
- All junctions with more than 50 reads have the same thickness
- Hovering the mouse over a junction will display the coverage

Question 4

- View → preferences → Alignments tab → Splice junction track options panel.
Choose Min flanking width=2 and Min junction coverage=10
- Right-click on the Refseq track → Expanded



Question 4 : Ensembl

www.ensembl.org : choose an Ensembl version on hg19 (GRCh37)

The screenshot shows the Ensembl genome browser interface. At the top, there is a search bar with the text "Search all categories" and "Search Human...". Below the search bar, there is a dropdown menu for "Human" and a "Go" button. The search results show "e.g. BRCA2 or 17:63973115-64437414 or osteoarthritis".

The main content area is divided into several sections:

- Genome assembly: GRCh38.p3 (GCA_000001405.18)**: This section includes links for "More information and statistics", "Download DNA sequence (FASTA)", "Convert your data to GRCh38 coordinates", and "Display your data in Ensembl". It also shows "Other assemblies" with a dropdown menu set to "GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart".
- Comparative genomics**: This section includes "What can I find? Homologues, gene trees, and whole genome alignments across multiple species.", "More about comparative analysis", and "Download alignments (EMF)".
- Regulation**: This section includes "What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and microarray annotations.", "More about the Ensembl regulatory build and microarray annotation", "Experimental data sources", and "Download all regulatory features (GFF)".
- Gene annotation**: This section includes "What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.", "More about models", "Download", "Update your models", and "View in archive site".
- Variation**: This section includes "What can I find? Disease and other variants", "More about variants", "Download", and "Variant Effect Predictor".

The "View in archive site" dropdown menu is open, showing a search bar and a list of links: "Help topics", "Frequently Asked Questions", "Video Tutorials", "Glossary", and "Contact HelpDesk".

The "View in archive site" section also displays a list of available archives for the selected gene set:


- Ensembl GRCh37: Full Feb 2014 archive with BLAST, VEP and BioMart
- Ensembl 80: May 2015 (GRCh38.p2) - patched/updated gene set Jan 2015
- Ensembl 79: Mar 2015 (GRCh38.p2)
- Ensembl 78: Dec 2014 (GRCh38)
- Ensembl 77: Oct 2014 (GRCh38) - patched/updated gene set Aug 2014
- Ensembl 76: Aug 2014 (GRCh38) - gene set updated Jul 2014
- Ensembl 75: Feb 2014 (GRCh37.p13)
- Ensembl 74: Dec 2013 (GRCh37.p13) - patched/updated gene set Sep 2013
- Ensembl 73: Sep 2013 (GRCh37.p12) - patched/updated gene set Jun 2013
- Ensembl 72: Jun 2013 (GRCh37.p11) - patched/updated gene set Apr 2013
- Ensembl 71: Apr 2013 (GRCh37.p10) - patched/updated gene set Feb 2013
- Ensembl 70: Jan 2013 (GRCh37.p8)
- Ensembl 69: Oct 2012 (GRCh37.p8) - patched/updated gene set Oct 2012
- Ensembl 68: Jul 2012 (GRCh37.p8) - patched/updated gene set May 2012
- Ensembl 67: May 2012 (GRCh37.p7) - patched/updated gene set Feb 2012
- Ensembl 54: May 2009 (NCBI 36) - patched/updated gene set Oct 2008

At the bottom of the page, there is a footer with the text "Ensembl release 81 - July 2015 © WTSI / EBI" and a link to "Permanent link - View in archive site".

Question 4 : Ensembl

Archive! Ensembl [BioMart](#) | [Tools](#) | [Downloads](#) | [Help & Documentation](#) | [Blog](#)





Human (GRCh37) ▾

 **Human**
Homo sapiens

Search all categories ▾

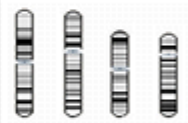
e.g. [BRCA2](#) or [6:133017695-133161157](#) or [osteoarthritis](#)

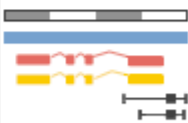
Genome assembly: GRCh37 (GCA_000001405.14)

-  [More information and statistics](#)
-  [Download DNA sequence \(FASTA\)](#)
-  [Convert your data to GRCh37 coordinates](#)
-  [Display your data in Archive EnsEMBL](#)

Other assemblies

Updated variation and regulation data on human assembly GRCh37 is available on our dedicated [GRCh37 archive site](#), along with a full suite of Ensembl web tools including VEP and BLAST.

 [View karyotype](#)

 [Example region](#)

Question 4 : Ensembl

Archive! Ensembl BioMart | Tools | Downloads | Help & Documentation | Blog Login/ Register

Human (GRCh37) Location: 19:11,683,475-11,691,801 Search all species...

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail
- Comparative Genomics
 - Alignments (image) (69)
 - Alignments (text) (69)
 - Region Comparison (79)
 - Synteny (17)

Chromosome 19: 11,683,475-11,691,801

Assembly exceptions
Chr. 19

Location: 19:11683475-11691801 Go

Gene: Go

Chromosome bands: 37_wav_GERP_ele..., Human_cDNAs (R..., Genes (GENCODE...

Constrained elements for 37 eutherian mammals EPO_LOW_COVERAGE

ZNF627-005 > processed transcript
ZNF627-004 > protein coding
ZNF627-006 > processed transcript

Contigs: AC009001.8 > AC020947.6 >

Genes (GENCODE...): < ACP5-002 protein coding, < ACP5-001 protein coding, < ACP5-003 protein coding, < ACP5-004 protein coding, < ACP5-005 protein coding, < ACP5-012 protein coding, < ACP5-013 retained intron, < ACP5-008 protein coding, < ACP5-004 retained intron, < ACP5-011 protein coding, < ACP5-010 retained intron, < ACP5-009 protein coding, < ACP5-007 protein coding

8.33 kb Forward strand

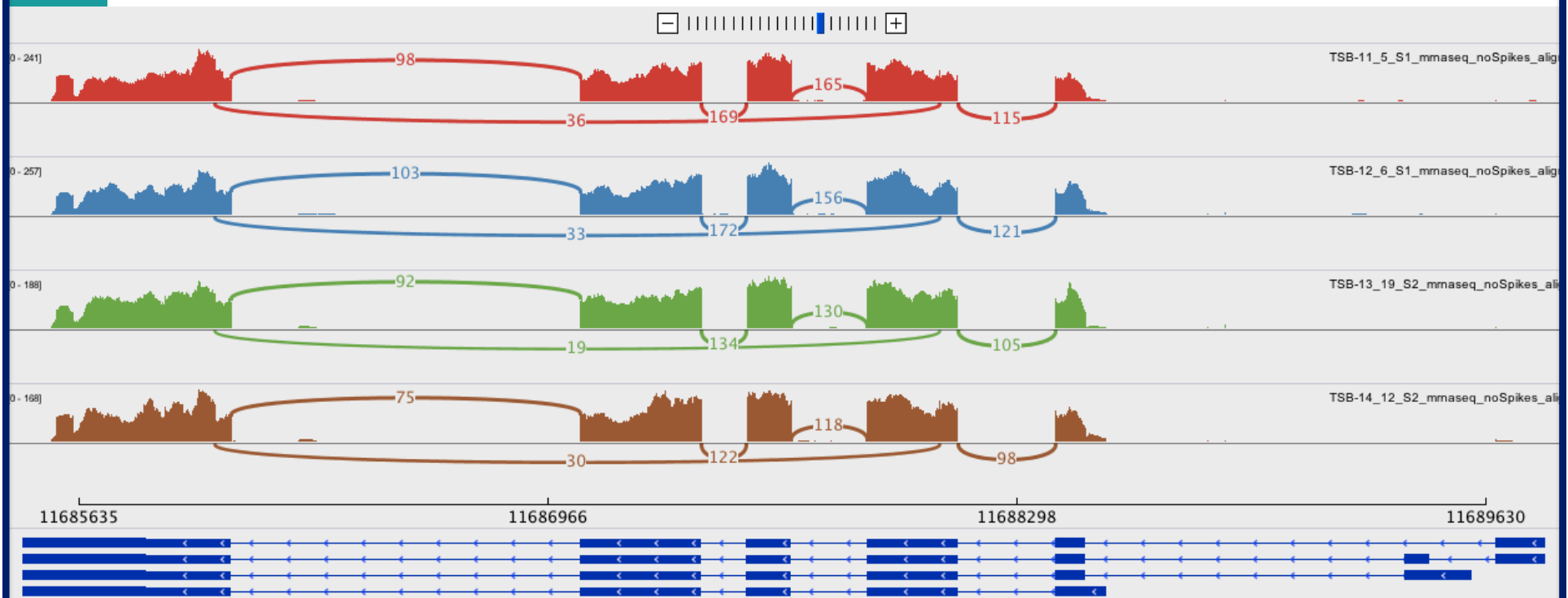
11.684Mb 11.685Mb 11.686Mb 11.687Mb 11.688Mb 11.689Mb 11.690Mb 11.691Mb

p13.2

This junction is in Ensembl annotations

Question 4

- Shashimi plot



➔ Very useful to quickly screen differentially spliced exons along genomic regions of interest